546-D GA1.13:FGMSD-81-4

REPORT BY THE

Comptroller General

OF THE UNITED STATES

COMPLETED

ORIGINAL

IRS Can Expand And Improve Computer Processing Of Information Returns

The Internal Revenue Service has made considerable progress in matching tax returns with information returns (such as W-2s and 1099s) filed by employers and organizations that pay dividends, interest, and other taxable income. However, IRS has not yet achieved the full matching program intended by the Congress.

IRS has sufficient computer capacity for a full matching program, but it will need additional funding and personnel to transcribe information returns and to follow up on unreported income.

IRS is planning to replace much of its computer equipment under a three-phase program that the agency estimates will cost $218 million over 8 years. It must concentrate now on preventing the use of outmoded, inefficient procedures with the new equipment. IRS should also pursue opportunities to automate the transcription of information returns, increase the filing of computer-readable information returns, increase computer program efficiency and equipment productivity, and improve master file design.

32

FGMSD-81-4

OCTOBER 20, 1980

Request for copies of GAO reports should be sent to:

U.S. General Accounting Office
Document Handling and Information
Services Facility
P.O. Box 6015
Gaithersburg, Md. 20760

Telephone (202) 275-6241

The first five copies of individual reports are free of charge. Additional copies of bound audit reports are $3.25 each. Additional copies of unbound report (i.e., letter reports) and most other publications are $1.00 each. There will be a 25% discount on all orders for 100 or more copies mailed to a single address. Sales orders must be prepaid on a cash, check, or money order basis. Check should be made out to the "Superintendent of Documents".

COMPTROLLER GENERAL OF THE UNITED STATES
WASHINGTON, D.C. 20548

B-200563

The Honorable Benjamin S. Rosenthal
Chairman, Subcommittee on Commerce,
Consumer and Monetary Affairs
Committee on Government Operations
House of Representatives

Dear Mr. Chairman:

Your letter of November 1, 1979, (app. I) asked us to (1) determine the progress made by the Internal Revenue Service toward a full information document matching program, (2) assess the adequacy of IRS computer capabilities for a full matching program, (3) determine the feasibility and costs of upgrading IRS computer capacity, (4) determine the general status of IRS computer capabilities, and (5) identify any other problem areas related to the IRS management and use of computers. On April 23, 1980, we briefed you on the results of our review, and at your request we are providing this written report.

SCOPE AND METHODOLOGY

We conducted our review at the IRS National Office, the IRS National Computer Center (NCC), the Andover (Massachusetts) and Fresno (California) Service Centers, and the Social Security Administration (SSA) in Baltimore. For tax year 1978, the most recent year for which information was available, 1/ we used IRS National Office reports to determine the number of information returns received and the number that entered the processing cycle. We attempted to verify and reconcile these reports with records at SSA and the Andover and Fresno Service Centers to the extent possible. We based our assessment of computer capacity at the National Computer Center on 1979 computer utilization statistics and related reports. Similar information on the computers at all 10 service centers was obtained from the IRS National Office and supplemented by our analyses at the Andover and Fresno Service Centers.

1/Processing 1 year's information returns takes about 3 years.

OVERVIEW

IRS has made considerable progress toward achieving a full document matching program, but still does not use millions of information returns. About 79 percent of the information returns received for tax year 1978 entered the processing cycle for the matching program. The IRS has sufficient computer capacity to process all the information returns, but additional funding and personnel are needed to transcribe the returns and follow up on unreported income. IRS plans to replace much of its computer equipment under a three-phase replacement program that IRS estimates will cost $218 million over 8 years. This replacement and conversion to machine independent software offers the potential to improve IRS computer operations, but IRS also has the opportunity to greatly increase the productivity of some of its present computers and thereby reduce costs and facilitate conversion to new systems.

PROGRESS TOWARD A FULL MATCHING PROGRAM

While making considerable progress toward the goal of a full document matching program, IRS has fallen somewhat short of that goal. IRS still does not use millions of information returns filed each year. In 1976, the Subcommittee on Commerce, Consumer and Monetary Affairs of the House Committee on Government Operations strongly recommended that IRS achieve a full document matching program by 1980. The Congress envisioned a full document matching program when it passed the Revenue Act of 1962. The act requires payers of wages, interest, dividends, and certain other taxable income to file information returns (see app. II) showing the amounts paid and the payees. Using computers, IRS matches some of these information returns to taxpayer returns to spot unreported income.

IRS is using in its information returns program about 79 percent of the 525 million returns filed for tax year 1978, compared with only 54 percent of tax year 1977 returns. This increase is largely attributable to the tax year 1978 implementation of the Combined Annual Wage Reporting Program that was established by Public Law 94-202. At least 108 million tax year 1978 information returns will not be used, and more returns were dropped from processing because the taxpayer identification numbers (TINs) were invalid, missing, or unverifiable. Appendix III shows the IRS estimates of the numbers of returns received and processed for tax years

1974 through 1978. Appendix IV provides a detailed discussion of reasons IRS does not use some information returns.

IRS HAS SUFFICIENT COMPUTER CAPACITY BUT NEEDS OTHER RESOURCES TO PROCESS AND USE ALL INFORMATION RETURNS

IRS has sufficient computer capacity to process and match all the information returns not currently being used--which would be about a 26-percent increase in the workload considering the approximately 108 million returns not processed for tax year 1978. Even with this increased workload, additional capacity is still available to process new types of information returns if reporting requirements were expanded. However, at some service centers--Austin (Texas), Fresno, and Ogden (Utah) in particular--the computers used in case followup are at or near capacity. The various segments of automated processing are discussed in appendix IV.

While IRS has more than enough computer resources to process all the information returns it receives, the agency says it would need additional staff to transcribe paper information returns not now used and to follow up on any additional tax cases identified through expansion of the matching program. Information returns received as paper documents cost more to process than those received on magnetic media because data on paper must be transcribed to computer-readable media. Therefore, IRS transcribes only a sample of the paper information returns received at the service centers. The bulk of unprocessed returns are paper returns. As shown in appendix V, IRS has estimated the costs and benefits of expanding the sample of paper returns from its current 15-percent level. For fiscal 1981, IRS proposed additional budget authority and positions to expand paper returns processing to a one-third sample, but this proposal exceeded budget ceilings imposed by Treasury.

Although transcribing paper information returns is primarily a manual process at IRS, there is the potential to expand processing through automation. For example, SSA uses optical character recognition (OCR) equipment to transcribe almost 60 percent of the paper wage information returns it receives. In 1978, IRS tested the feasibility of using OCR equipment to process information returns. The results were not good because the forms tested were not standardized or designed for the equipment. IRS has not pursued the automation of this processing but we believe the agency should, perhaps with SSA. IRS also should encourage payers

to file returns on computer-readable media and to improve the quality of their submissions to both SSA (W-2s) and to IRS (1099s and 1087s).

IRS PLANS TO REPLACE ITS COMPUTERS

IRS plans to replace much of the computer equipment used in tax processing under a three-phase equipment replacement program that will cost about $218 million for hardware and software over 8 years. In estimating workload requirements for the replacement system at the service centers, IRS did not project the requirements for individual programs like the information returns program. Instead, the IRS estimated an overall 8 percent annual compounded rate of growth in workload, including 5 to 6 percent annual growth for unforeseen new requirements. Without a detailed analysis of all processing requirements, we could not determine if the overall growth factor provides sufficient growth for a full document matching program.

Under phase I of the equipment replacement program, IRS will replace the H-2050A and CDC 3500 computers with a single system at all 10 service centers. Plans call for the replacement system to be partially operational at 1 center by October 1982 and completely operational at all 10 centers by January 1985. IRS issued a request for proposals for phase I in February 1980. Phase II will replace the nine computers at the National Computer Center between 1984 and 1986, and phase III will automate the microfilm research activities at the service centers from 1984 through 1986. Since IRS has not yet performed feasibility studies for these systems, we could not assess or identify their potential impact on information returns processing.

GENERAL STATUS OF CURRENT IRS COMPUTERS

The current IRS computer systems generally consist of old equipment and software that is slower, more expensive to operate, and less flexible than systems currently available. Application computer programs are primarily (98 percent) programmed in machine-dependent assembly language which makes them more expensive to maintain and change and difficult to shift to other makes or newer models of computer systems. For example, the service centers cannot easily shift workload from the heavily burdened CDC 3500 computers to the underutilized H-2050A computers. Also, the National Computer Center cannot easily convert to more efficient operating

systems because the computer programs have been customized to run under operating systems that are now obsolete.

Although hardware maintenance is not yet a significant problem on the older systems, the lack of some modern hardware features does create operational problems. For example, the H-2050A computers do not have the capability to process the high density magnetic tapes now in common use. Thus, about 20 percent of the information returns received on magnetic tape at Fresno must be sent to the National Computer Center to be converted to lower density tape and then sent back to the service center for processing.

As previously noted, at some service centers the workload on the CDC 3500 has reached or is near capacity. These systems cannot be further modularly upgraded to reflect differences in the workload distribution among centers. We did not perform a detailed review of the feasibility of redistributing the workload among centers, but a practical approach appears to be a site-by-site analysis and sizing of the planned replacement systems to the workload of each center.

OPPORTUNITIES TO IMPROVE COMPUTER PRODUCTIVITY

Inefficient design of the information returns processing system and mismatches between job requirements and computer resources are limiting the productivity of the IRS computers. IRS could improve computer productivity and thereby increase the capacity and output of its systems by improving the information returns system design and by better matching of job requirements and computer resources.

Need improved system design

During 1979 an IRS team made a comprehensive study of the information returns program. The draft study report 1/ recommended legislative and many other changes to information reporting requirements and a streamlined system design for processing the information. Some of the recommendations could substantially improve use of the IRS computers. Several

---

1/"System Design Recommendations for an Improved Information Returns Program," August 1979.

of the underlying problems, similar to those identified by the study team, were addressed in our June 1979 report "IRS Can Better Plan for and Control its ADP Resources" (GGD-79-48, June 18, 1979).

The study group reported that file size and processing time could be reduced by about one-half throughout the system. The current system design was described as the result of a "series of tack-ons to computer programs originally designed in the mid-1960s," and the report indicates that the stability of the system is decaying while maintenance costs and other resource requirements are increasing. Broad application design recommendations included consolidating computer programs using standardized program routines, redesigning record and file structures, using simple run control techniques, and improving the conversion of data from paper documents.

The study group suggested implementing the above recommendations before or concurrent with the phase II equipment replacement at the National Computer Center. The recommendations are currently being reviewed by IRS management. We believe this redesign and development in higher level languages for the information returns program would be more appropriately done prior to the phase II acquisition.

The impending phase II acquisition is planned as a fully competitive replacement which will involve conversion of over 2,000 application programs written in machine-dependent language, as well as transition of all personnel to operating with new hardware and systems software. An effort of this magnitude could take 3 to 10 years to complete. IRS has sufficient capacity for its needs for the foreseeable future.

In view of the upcoming phase II acquisition and the present adequate capacity for up to the next 5 years at the National Computer Center, we believe IRS has an immediate opportunity to plan, redesign, and develop in higher level language, its individual and business master file system. Improved system design for the information returns program would, of course, be part of this. The master file redesign would incorporate (1) a new file design, (2) higher level language applications, and (3) expanded functional uses and capabilities such as matching information returns to business tax returns.

If an overall redesign is not done, we believe IRS will continue to (1) perpetuate limitations on functional use of the vital master file, (2) experience lengthy production cycles, (3) be locked into archaic file and application programs, and (4) foster more noncompetitive acquisitions. At the time of our review the IRS had not yet begun the feasibility study for the phase II replacement.

Need to better match job requirements and computer resources

IRS has configured the nine National Computer Center computers in a way that allows them to be used interchangeably but sacrifices productivity. The fixed rules for using the computers do not provide for matching the needs of individual jobs to the computer resources, because all jobs are assumed to require the same (maximum) resources. As a result, system capabilities are grossly underutilized. By using modern automated scheduling techniques and resource requirement statements for each job, IRS could increase substantially the available capacity to accommodate additional workloads, produce more, and reduce costs. For example, by increasing the productivity of the two IBM 370 systems and the new Itel system, IRS could probably increase services and provide for more timely output, and/or possibly eliminate some or all of the older IBM 360s.

In our June 1979 report (GGD-79-48, June 18, 1979), we identified similar system inefficiencies and recommended an improved computer performance management program. IRS accepted our recommendations. The continuation of the type of problem just cited indicates that IRS should increase its emphasis on computer performance management and strengthen the role of the ADP performance evaluation staff in this management. Because IRS plans to replace the National Computer Center computers beginning in 1984, we believe a concerted effort is needed now to increase the productivity of these systems rather than "roll-over" old inefficient procedures to the new equipment. We believe that action by senior management and participation by the ADP performance evaluation staff is needed if IRS is to fully and effectively obtain the processing capabilities of the present and new computers.

CONCLUSIONS AND RECOMMENDATIONS

IRS has made considerable progress toward full document matching envisioned by the Committee on Government Operations

in its 1976 report, but has fallen somewhat short of reaching that goal. Computer capacity is not a constraint to full document matching, but staffing and funds are. An opportunity exists to expand the use of information returns by automating transcription of paper returns, by reducing the need for transcription by obtaining returns in computer-readable form, and by cooperating with SSA in both aspects. IRS' use of its data processing systems could be substantially improved by improving the design of the information returns system and by better matching job requirements to computer resources.

We recommend that the Commissioner of Internal Revenue:

--Investigate the use of OCR equipment to transcribe non-wage paper information returns, possibly through equipment sharing with SSA.

--In cooperation with SSA, seek ways to increase the filing of information returns on computer-readable media.

--Redesign the information returns processing system to improve efficiency along the lines recommended in the IRS draft report entitled "System Design Recommendations for an Improved Information Returns Program," August 1979.

--Require the ADP performance evaluation staff to develop a comprehensive resource-need-based workload scheduling and management system to improve computer productivity at the National Computer Center.

--Do a feasibility and cost-benefit study of a redesign, and development in higher level language, of the business and individual master files and related applications presently written in machine-dependent languages.

IRS COMMENTS AND OUR EVALUATION

In his September 5, 1980, letter (app. VI) the IRS Commissioner generally concurred with all of our recommendations and, in most cases, indicated that action to comply with our recommendations had been taken or would be. In some cases, however, no timetable was provided and we are somewhat concerned about the timeliness of the action.

Regarding our recommendation that IRS begin a redesign of the master files, the Commissioner replied that IRS is focusing its efforts on equipment replacement and conversion of programs to a higher level language. He stated that staff resources and skills are not available to simultaneously redesign the master files and he provided no target date for this redesign.

We believe IRS should seek the resources needed to begin redesign of the master files now. The master file design is a critical segment of the entire tax processing system. The design and use of the master files are keys to the effective use of most tax data and thus to the overall effectiveness of IRS. The resources needed to accomplish this planning and design work are modest and should not affect the current system replacement efforts. Thus, we believe the redesign effort should begin now to ensure that the computer systems and software obtained in the replacement program are the most appropriate to IRS needs and to ensure the new equipment will enable the most effective and efficient use of the master files.

As arranged with your office, we will publish this report as soon as possible after your hearings on October 1, 1980. We will then send copies to the Director of the Office of Management and Budget, the Secretary of the Treasury, the Commissioner of Internal Revenue, and to concerned congressional committees. We will also make copies available to the press and the public.

Sincerely yours,

Comptroller General
of the United States

LIST OF APPENDIXES

APPENDIX		Page
I	November 1, 1979, letter from the Chairman, Subcommittee on Commerce, Consumer and Monetary Affairs, House Government Operations Committee	11
II	Types of payments subject to information reporting	13
III	Information returns received and processed for tax years 1974 through 1978	14
IV	Many information returns not used but computer capacity exists for additional processing	15
	Why returns are not used	15
	IRS has sufficient computer capacity to process all information returns	18
V	IRS cost-benefit analysis for expanding the sample of paper information returns	27
VI	September 5, 1980, letter from the Commissioner of Internal Revenue	28

APPENDIX I

BENJAMIN S. ROSENTHAL, N.Y., CHAIRMAN
ROBERT T. MATSUI, CALIF.
EUGENE V. ATKINSON, PA.
FERNAND J. ST GERMAIN, R.I.
JOHN CONYERS, JR., MICH.
ELLIOTT H. LEVITAS, GA.

LYLE WILLIAMS, OHIO
[illegible]

NINETY-SIXTH CONGRESS

Congress of the United States

House of Representatives

COMMERCE, CONSUMER, AND MONETARY AFFAIRS
SUBCOMMITTEE
OF THE
COMMITTEE ON GOVERNMENT OPERATIONS

RAYBURN HOUSE OFFICE BUILDING, ROOM B-377
WASHINGTON, D.C. 20515

November 1, 1979

Hon. Elmer B. Staats
Comptroller General
General Accounting Office
441 G Street, N.W.
Washington, D. C. 20548

Dear Mr. Comptroller General:

On September 23, 1976, the House Government Operations Committee issued a report prepared by the Subcommittee on Commerce, Consumer and Monetary Affairs entitled "Effectiveness of the Internal Revenue Services Income Information Document Matching Program." In this report the committee recommended that "IRS should move to a full information document matching program by fiscal year 1980. The IRS should keep the committee fully informed of its progress in meeting this goal." It was also recommended that IRS make this a high priority item and "seek sufficient funds from Congress for this purpose."

The subcommittee is continuing its investigation into IRS' document matching program and therefore requests the General Accounting Office to initiate a study into the IRS' resource capability for moving to a full document matching program. This study should be so oriented as to ultimately place GAO in the position to be able to testify before the subcommittee.

Those points of most immediate concern to this subcommittee are:

1. What would be the additional workload (in general or relative terms) to implement a full document matching program? Where does IRS stand in terms of implementing such a program?

2. Does the IRS have the computer hardware capability to absorb the additional computer runs necessary to perform a full document matching program?

3. Can the IRS update its current computer capacity and if so at what cost? In other words, what are the hardware and software costs involved?

4. What is the general status of IRS' computer capabilities?

   a. Do they use machine dependent language in violation of Federal Statutes standards and regulations (P.L. 89-306)?

APPENDIX I

b. Is their hardware antiquated to the point of having support problems or being inefficient? What kind of operational problems do they have with their hardware?

c. What are the distributional workload problems among service centers? Can these problems be worked out? If so, how?

5. Any other problem areas related to the efficient and effective management and use of ADP in support of the IRS mission.

As IRS has limited computer capability at this point in time, I urge GAO to expedite this request. A phased approach with an initial staff report due in late February with interim progress reports as appropriate will facilitate the subcommittee's time schedule.

If you have any questions concerning this request, please contact Herschel F. Clesner, subcommittee chief counsel, at 225-4407.

Thank you for your cooperation and assistance.

Sincerely,

Benjamin S. Rosenthal
Chairman

BSR:cb

TYPES OF PAYMENTS SUBJECT TO INFORMATION REPORTING

Type of Payment (note a)	Information Returns
Individual wages	Form W-2
Pensions and annuities	Form W-2P
Gambling winnings	Form W-2G
Interest	Forms 1099, 1087 INT
Dividends	Forms 1099, 1087 DIV
Miscellaneous business payments	Forms 1099, 1087 MISC
Insurance plan payments to health care suppliers	Forms 1099, 1087 MED
Original issue discounts on certain indebtedness	Forms 1099, 1087 OID
Compensation for services by nonemployees	Form 1099 NEC
Compensation to certain fishing boat crew members	Form 1099 F
Taxable distributions from cooperatives	Form 1099 PATR
Lump sum distributions from profit-sharing and retirement plans	Form 1099 R
Bearer certificates of deposit	Form 1099 BCD
Unemployment compensation	Form 1099 UC
Liquidation distributions	Form 1099 L
Individual retirement accounts	Form 5489
Distributions of earnings of small business corporations, partnerships, and fiduciaries	Schedule K-1
Foreign income	-
Agricultural subsidies	-

a/Under certain circumstances information returns may not be required for these payments. For example, some payments under certain amounts are exempt.

INFORMATION RETURNS RECEIVED AND PROCESSED
TAX YEARS 1974 - 1978

	Tax year 1974		Tax year 1975		Tax year 1976		Tax year 1977		Tax year 1978 (note a)	
	Number	Percentage of total received	Number	Percentage of total received	Number	Percentage of total received	Number	Percentage of total received	Number	Percentage of total received
	(millions)									
Received:										
on magnetic media	186	41	217	46	248	52	276	55	316	60
on paper	265	59	255	54	233	48	228	45	209	40
Total	451	100	472	100	481	100	504	100	525	100
Processed to NCC:										
from magnetic media	163	88%	188	87%	207	83%	232	84%	278	88%
from paper	17	6%	36	14%	36	15%	39	17%	139	66%
Total	180	40%	224	47%	243	50%	271	54%	417	79%
Matched	158	35%	201	42%	228	47%	253	50%	397	76%

a/Beginning with tax year 1978, SSA received wage information returns and tried to process all these returns to NCC. The figures shown represent the combined number of returns received by IRS and SSA. The total received includes an estimated 24 million returns received by SSA but not processed to IRS in time for use in the matching program.

APPENDIX III

APPENDIX IV

MANY INFORMATION RETURNS NOT USED BUT COMPUTER CAPACITY EXISTS FOR ADDITIONAL PROCESSING

WHY RETURNS ARE NOT USED

At least 108 million information returns filed for tax year 1978 will not be used in the matching program. They include paper returns not transcribed to magnetic media, business information returns, wage returns not processed in time by SSA, and other returns as described below.

Statistical sample of non-wage returns

Since returns filed on paper are more expensive to process than returns filed on magnetic media, IRS processes only a sample of most types of non-wage paper returns. The sampling plan is based on the first letter of the payee's surname and is designed to yield a usable 15-percent sample; more than 15 percent are selected to allow for drop-outs. For tax year 1978, IRS service centers processed to NCC about 22 percent of the paper returns they received; about 57 million were not processed.

As shown in appendix V, IRS has estimated the costs and benefits of expanding the paper sample to various levels. For fiscal 1981, IRS proposed additional budget authority and positions to expand the sample to one-third of the paper returns, but this proposed exceeded budget ceilings imposed by Treasury. The IRS 1981 budget is now pending before the Congress.

Business returns are not matched

IRS matches information returns only to tax returns filed by individuals; therefore, an estimated 34 million returns pertaining to businesses are not processed. Problems matching information returns to business tax returns include: differences in reporting (e.g., calendar vs. fiscal year and cash vs. accrual basis); difficulty identifying payees (e.g., parent vs. subsidiary and business trusts vs. personal trusts); and tax return master files that do not contain line items needed for matching to information returns.

An IRS task force convened in 1976 determined that a business return matching program is feasible. IRS is currently testing methods to implement such a program, using a 1-percent sample of business taxpayers. The study will be completed in

APPENDIX IV

December 1980, and, if approved, the system could become operational for tax year 1982.

Wage returns not processed in time by SSA

About 24 million tax year 1978 wage information returns received by SSA were not processed to IRS in time for use in the matching program. During this first year of combined annual wage reporting, program delays occurred due to misunderstandings between SSA and IRS, the poor quality of some returns received, and operational problems at SSA, as described below.

--SSA had to manually check and correct about 12 million processed returns because of a misunderstanding between IRS and SSA on editing requirements.

--About 50 percent of magnetic media reports could not be read on the first try, thereby requiring contact with employers. Problems included variance from requested formats and returns that did not balance to summary statements.

--Some employers did not follow instructions to separately batch all returns without a taxpayer identification number.

--About 42 percent of paper returns were not suitable for scanning by optical character recognition equipment. SSA, therefore, had to transcribe the returns to magnetic media using key data entry equipment. Of the returns not scanned, about 41 percent used unacceptable ink, 36 percent used nonstandard print, and 10 percent used nonstandard forms. 1/

--Upgrading of the key data entry equipment was not completed when the returns began to arrive in January 1979.

--SSA had to re-microfilm over 3.5 million returns to meet IRS image clarity standards. Also, a system to monitor and control batch and document locator numbers was not implemented until 3 months after microfilming had begun. (Since SSA does not transcribe payees' addresses--which IRS uses to research missing or

---

1/The percentages are based on an SSA sample of tax year 1979 returns, which SSA officials believe are representative of the 1978 returns.

invalid TINs--IRS uses the microfilm copies to obtain addresses on returns requiring research.)

SSA anticipates more timely processing of tax year 1979 returns, but SSA and IRS have agreed to a contingency plan to be implemented in the event of delays. Of those returns not processed by SSA, IRS will process all returns within the statistical sample used for non-wage returns filed on paper.

Other reasons returns are not used

For tax year 1977, at least 5.6 million returns were dropped from processing because of invalid or missing TINs. For returns within a statistical sample, 1/ IRS tries to determine the correct TIN by using computers to compare payee name and address keys to NCC records, then manually screening lists of possible matches. IRS has programs to notify payers of required corrections to TINs and to assess penalties to payers that file returns without TINs.

Only about half the estimated 600,000 foreign information returns received for tax year 1978 were used. The remaining returns were either received too late or lacked required information. Also, some countries file no returns or returns that do not include all income. IRS has contacted its tax treaty partners to improve the quality and timeliness of their returns and to expand reporting to types of income presently excluded.

Because of budget constraints, beginning with tax year 1976 the Bureau of Public Debt stopped providing IRS with bondholder names on tape returns for Series E savings bonds. Since IRS uses both the name and TIN in the matching process, IRS did not use the tax year 1976 and 1977 returns--an estimated 2 million returns for 1977 alone. For tax year 1978, IRS developed a method to obtain the names by matching TINs to SSA records.

Returns for liquidating distributions (Form 1099 L) do not separately identify nontaxable distributions or capital or short-term gains and losses. Because of the difficulty

---

1/The sampling method used is the same as that used to select non-wage returns filed on paper, as discussed on page 14. However, the TIN-correction sample applies to all returns--wage and non-wage; paper and magnetic media.

APPENDIX IV

in matching the composite amount to tax returns, IRS has not used these returns. However, IRS plans to try matching these returns beginning with tax year 1980.

Other returns are not processed either because payment amounts are below IRS criteria or because the returns are not readable. According to IRS each of these factors account for about 2 percent of all returns.

IRS HAS SUFFICIENT COMPUTER CAPACITY TO PROCESS ALL INFORMATION RETURNS

IRS has sufficient computer capacity to process and match all the information returns not currently being used. Considering the approximate 108 million returns not processed for tax year 1978, this would be about a 26-percent increase in the information returns workload. Even considering this increased workload, additional capacity is available to process new types of information returns if reporting requirements are expanded. However, at some service centers the computers used in case followup are at or near capacity.

Processing 1 year's information returns takes about 3 years, as shown below.

Processing phase	Processing activity	Calendar year in which activity was performed for tax year 1977 information returns
I	Receive, transcribe, and reformat information returns; validate and correct TINs	1978
II	Cross-reference spouses' returns, extract tax return data from master files, and prepare files for matching	1979
III	Match information return and tax return files and follow up on unreported income	1979
IV	Prepare statistical reports on program results	1980

The various segments of automated processing and the capacity the computers used in each segment are separately discussed below.

Data transcription

Each service center transcribes data from paper to magnetic tape using a Direct Data Entry System (DDES) consisting of over 400 key entry terminals and various peripheral devices linked to dual GE 4020 computers.1/ The system was installed in 1969, and IRS has replaced and augmented portions of the DDES to handle growth and new applications and to improve efficiency.

The information returns workload on the DDES has sharply declined. Returns filed on magnetic media do not require data transcription and account for about 79 percent of all tax year 1978 returns filed at service centers, compared to about 41 percent in tax year 1974. As shown below, with the implementation of the combined annual wage reporting program, the number of paper returns received and processed at the IRS service centers declined.

	Tax year 1977	Tax year 1978
	----- millions -----	
Paper returns received at IRS service centers	228	73
Paper returns processed at IRS service centers	39	16

If all paper returns received in tax year 1978 were processed, the additional DDES workload would have been about 57 million returns.

1/The computers were manufactured by General Electric, but in 1971, the maintenance of GE computers was taken over by Honeywell.

APPENDIX IV

Based on our review at the Andover and Fresno Service Centers, 1/ the DDES equipment has sufficient capacity for transcribing all paper returns. For example, at Andover we estimate 121,000 DDES terminal hours would be required to process the 4.7 million tax year 1978 paper returns not processed. The Andover DDES has a total capacity of about 1.6 million terminal hours for the January to June peak workload period, based on 448 terminals staffed 20 hours per day, 7 days per week. For January to June 1980, Andover expects to staff the DDES for 490,000 hours, leaving sufficient equipment availability to process the paper information returns. Andover and Fresno IRS officials agreed there is sufficient DDES equipment capacity to process all the paper returns, but they voiced concern about the feasibility of hiring and retaining sufficient personnel.

In 1978, IRS and two OCR equipment vendors tested the feasibility of using OCR to process various Form 1099 paper information returns. It represented a "worst-case" test, since the returns used were on forms not designed for OCR use and were prepared by payers under current non-OCR procedures. IRS concluded that to be scannable, the forms would have to be standardized and redesigned for OCR use, payers would have to be instructed to follow new requirements for data placement, and print quality would have to be improved. IRS decided not to pursue these changes pending evaluation of SSA experience in using OCR to process wage information returns.

Although the DDES has sufficient capacity to process all paper information returns, we believe that IRS should continue to explore alternatives. SSA has a full year's experience in using OCR equipment to process paper W-2 forms, and a cooperative IRS-SSA program could offer the potential to expand OCR processing to other types of returns. Another alternative is to centralize transcription of paper returns at those service centers with sufficient labor supply. Finally, IRS should continue to seek ways to further increase filing of returns on magnetic media.

1/The Andover and Fresno Service Centers receive about 8 percent and 6 percent, respectively, of all paper returns received at the 10 service centers.

File editing, reformating, and related processing

Each service center uses a Honeywell H-2050A computer to edit, reformat, and perform other processing on tape files of information returns to be sent to NCC, returns with invalid or missing TINs received from NCC, and tax cases identified through the matching process at NCC. During 1979, this processing accounted for a relatively small portion of the H-2050A workload--from 4.8 to 10.2 percent of the monthly utilization at all service centers combined.

In a full matching program, the additional workload on the H-2050A would include input processing for paper returns not currently used and TIN-correction processing for those returns and others not now included in the TIN-correction sample described on page 17. As shown below, more than sufficient H-2050A capacity is available during the months it would be needed for this processing.

	Andover		Fresno	
	Needed (note a)	Available (note b)	Needed (note a)	Available (note b)
	multiprogramming hours			
Input processing (Feb.-June)	298	2,377	403	894
TIN-correction and related processing (note c) (Aug.-Dec.)	776	1,857	534	1,221

a/Based on tax year 1978 processing times. Assumes that processing time increases in direct proportion to the number of returns processed.

b/Estimated by multiplying idle hours per month by the average number of jobs in multiprogramming during the month.

c/Assumes that the proportion of returns with invalid TINs remains constant.

IRS officials agreed that the H-2050A computers have sufficient capacity available to process to NCC all the information returns received at the service centers. The additional workload on the H-2050A to process tax cases identified through the matching program is separately discussed on page 25.

Microfilm research on wage returns

As noted on page 17, SSA sends IRS microfilm copies of wage returns, and IRS obtains payee addresses from the microfilm as a preliminary step in correcting invalid or missing TINs for returns filed on paper. This "pre-TIN" processing is centralized at the Andover Service Center. For returns filed on magnetic media, SSA provides NCC a separate magnetic tape file of payee addresses for use in TIN-correction.

Andover personnel used a manual system for pre-TIN microfilm research on tax year 1978 returns, but a recently installed Wage Information Retrieval System (WIRS) will be used to help process tax year 1979 returns. WIRS consists of a Four Phase minicomputer and storage devices linked to 16 data entry/display terminals and 16 microfilm reader/printers. WIRS operators use the system to search the microfilm and transcribe payee addresses to magnetic tape. We were told that with planned system modifications WIRS will have sufficient capacity to process about 550,000 returns using the system one shift per day, 5 days per week during a May through September pre-TIN processing schedule.

Andover performed pre-TIN processing on about 110,000 tax year 1978 returns, but we estimate about 3.6 million invalid--or no-TIN--returns were received by SSA. The returns not sent to Andover include returns processed too late by SSA, returns not included in the TIN-correction sample, and returns with payment amounts below processing criteria. Although WIRS has sufficient capacity to process within existing schedules about five times the tax year 1978 returns processed, it could not process all 3.6 million returns unless schedules were lengthened and shifts added. We believe IRS should evaluate (1) the costs and benefits of expanding the pre-TIN processing and (2) other alternatives such as strengthening programs to promote and enforce correct reporting of TINs.

Document matching

Information returns are matched to tax returns using computers at NCC, which maintains the tax return master files. Before the matching, NCC uses master files to check the validity of TINs, identify potential TIN-corrections, cross-index spouses' returns, and extract tax return data to be matched to the information returns. This processing accounted for only about 7 percent of the NCC 1979 workload.

NCC uses nine large-scale computer systems and rates their capacity in terms of IBM 360/65 "equivalents." As shown below, the nine computer systems provide NCC with 15 IBM 360/65 equivalents, or 124,800 hours of processing capacity per year based on 20 eight-hour processing shifts per week.

		Processing capacity	
Computer model	Quantity	IBM 360/65 equivalents	Hours
IBM 360/65	6	6.0	49,920
IBM 370/165	1	3.4	28,288
IBM 370/168	1	3.6	29,952
Itel AS/6	1	2.0	16,640
Total	9	15.0	124,800

IRS plans to upgrade the Itel AS/6 to three IBM 360/65 equivalents in 1982, thereby increasing total processing capacity to 133,120 hours per year.

Under an expanded document matching program, the additional workload on NCC computers would include processing the returns not currently used (about a 26-percent workload increase, as discussed on p. 18) and identifying potential TINs for returns not now included in the TIN-correction sample (about triple the current TIN-correction workload). As shown below, NCC computers have more than sufficient capacity to process this additional workload.

APPENDIX IV

	IBM 360/65 equivalent hours		
	1980	1981	1982
Total computer capacity	124,800	124,800	133,120
Forcasted workload for information returns processing at current program levels (note a)	9,970	10,220	10,480
Additional processing for a 26-percent increase in information returns	2,590	2,660	2,720
Additional processing for TIN-correction (note b)	600	620	640
Forecasted workload for all other applications (note a)	99,480	107,530	116,980
Total forecasted workload	112,640	121,030	130,820
Surplus capacity	12,160	3,770	2,300

a/Estimates provided by IRS.

b/Estimates equal three times tax year 1978 processing time plus workload growth.

IRS computer utilization statistics also show that the NCC computers are used in a way that greatly limits their productivity. As shown in the table below, the six IBM 360/65 systems were active (at least one program was loaded) an average of 71 to 72 percent of the time they were available, and the processors were busy only about 53 to 56 percent of the time the systems were active. The IBM 370/165 processor was more busy, but never more than about 79 percent of the time it was active. Although the IBM 370/165 is capable of running a balanced mix of 10 or more jobs concurrently, on the average it runs 2.8 jobs or less. (Comparable data was not available for the IBM 370/168, and the Itel AS/6 was not operational until March 1980.)

Computer	Percent of available time system was active (note a)			Percent of active time processor was busy (note a)		
	Low month	14-month average	High month	Low month	14-month average	High month
IBM 360/65	59	71	80	45	54	62
IBM 360/65	57	71	86	47	54	61
IBM 360/65	60	71	84	50	54	58
IBM 360/65	60	72	81	45	56	62
IBM 360/65	50	72	87	45	56	64
IBM 360/65	61	71	82	41	54	60
IBM 370/165	80	86	93	57	73	79
IBM 370/168	62	80	91	(b)	(b)	(b)

a/Based on IRS monthly equipment utilization trend reports. Due to time constraints we did not analyze the reliability of the source data used in those reports.

b/Information not available.

IRS officials agreed that the existing and planned computer capacity at NCC is sufficient for an expanded matching program. We believe that by improving computer utilization, the processing capacity at NCC can be increased.

Case followup

NCC sends the service centers tape files of potential tax cases identified through the document matching program. As discussed below, each service center uses its H-2050A computer and a Control Data Corporation CDC-3500 computer to help process the cases.

The H-2050A computers are used to prepare cases for printing and manual screening, prepare notices to taxpayers, and prepare case files. Although we could not readily estimate the number of cases that might result from a full matching program, as shown below there was sufficient capacity on the Andover and Fresno H-2050A computers during 1979 to greatly expand case file processing.

APPENDIX IV

	Andover	Fresno
	--multiprogramming hours--	
Capacity used for case processing in 1979	381	605
Unused capacity in 1979 (note a)	4,046	1,808

a/Unused capacity after deducting time needed for additional input processing and TIN-correction shown on p. 21.

The CDC-3500 computers support an Integrated Data Retrieval System (IDRS) used to retrieve and update information on taxpayer accounts, assign and control cases, prepare computer-generated correspondence, and perform various other functions. Each service center has about 400 remote terminals linked to its CDC-3500 for real-time processing during the day. During night and weekend shifts the CDC-3500 is used in a batch processing mode to maintain the IDRS files.

IRS estimates that the maximum real-time processing capacity of the CDC-3500 is 30,000 transactions per hour, based on a maximum acceptable response time of 6 seconds on 95 percent of all transactions. IRS analyses show that workload will reach capacity at three service centers in 1982. Also, at the Fresno Service Center, which has the heaviest CDC-3500 workload, on several occasions time required for weekend batch processing has exceeded time available.

We could not determine from IRS records what portion of CDC-3500 utilization is attributable to processing cases identified through the matching program, nor could we estimate the additional workload on the CDC-3500 from an expanded document matching program. Depending on the number of additional cases identified, workload saturation on the CDC-3500 could limit use of this system for case followup at some centers unless other processing is reduced or case followup criteria is adjusted to focus on those cases with the greatest potential payoff. However, as discussed on page 4, beginning in August 1982 IRS plans to replace the CDC-3500 and H-2050A computers with a single system at each service center.

IRS COST-BENEFIT ANALYSIS FOR EXPANDING THE SAMPLE OF PAPER INFORMATION RETURNS (note a)

Tax Year 1980 Estimates for Processing All Magnetic Media Returns Plus:

	15 percent paper	33-1/3 percent paper	75 percent paper	100 percent paper (note b)
	----------(millions)----------			
Gross benefit (revenue plus refunds)	$967	$1,025	$1,125	$1,180
Net yield (revenue less refunds)	483	515	565	590
Program cost	85	102	118	125
Gross benefit to cost ratio	11.4 to 1	10.0 to 1	9.5 to 1	9.4 to 1
Net yield to cost ratio	5.7 to 1	5.0 to 1	4.8 to 1	4.7 to 1

a/These estimates were prepared by IRS in September 1979 for the fiscal 1981 budget presentation. We did not evaluate the assumptions or data underlying these estimates.

b/The program costs and benefits at this level reflect transcription of all paper information returns, but not all would be matched to tax returns. The information returns not matched would include returns for amounts below a tolerance level and returns with invalid or missing TINs which could not be corrected.

APPENDIX VI

COMMISSIONER OF INTERNAL REVENUE

Washington, DC 20224

SEP 5 1980

Mr. William J. Anderson
Director
General Government Division
General Accounting Office
Washington, D.C. 20548

Dear Mr. Anderson:

We appreciate the opportunity to review your draft report entitled "Internal Revenue Service Computer Capabilities for Matching Information Returns to Tax Returns." Our comments on the recommendations and other items in the report are as follows:

Recommendation 1:

Investigate the use of OCR equipment to transcribe paper information returns, possibly through equipment sharing with the Social Security Administration (SSA).

Response:

We concur that the use of OCR equipment to transcribe paper information returns should be investigated to determine if this would be a cost-beneficial alternative.

Recommendation 2:

In cooperation with SSA, seek ways to increase the filing of information returns on computer readable media.

Response:

To expand the field of recruitable magnetic media filers for both IRS and SSA, the two agencies are surveying paper filers to determine the most prevalent types of computer equipment now available to them. With this information the two agencies can then determine what computer conversion equipment would be most feasible to expand the range of equipment acceptable for magnetic media filing and share such "translators" at a central location. IRS and SSA are currently sharing conversion equipment located at the Austin Service Center.

Recommendation 3:

Redesign the information returns processing system to improve efficiency along the lines recommended in the report entitled "System Design Recommendations for an Improved Information Returns Program (IRP)" dated August 1979.

Response:

Data Services is implementing the recommendations in the August 1979 study which relate to rewriting the computer programs used to process the information returns. These actions will improve efficiency. The rewriting of the computer programs has already begun as part of the conversion to the Service Center Replacement System. Other recommended actions contained in the August 1979 study are being considered by the newly established Internal Revenue Service IRP Control Group and IRP Coordinator. Items selected for implementation will be submitted to the ADP Policy/Resource Board for approval.

Recommendation 4:

Require the ADP Performance Evaluation Staff to develop a comprehensive resource need based workload scheduling and management system to improve computer productivity at the National Computer Center (NCC).

Response:

In May 1980, the ADP Performance Evaluation Staff completed a Workload Characterization Report on NCC. (A draft copy of the report was provided to GAO.) On the basis of some of the recommendations contained in the May 1980 report, an action plan has been prepared for designing and implementing an improved performance system. This system will be operational by the end of 1980. Emphasis will be placed on making tools and techniques available to NCC to improve short range scheduling and provide a series of reports for various levels of management. The feasibility of acquiring an automated scheduling system will also be considered. In addition, a Multiple Virtual Storage (MVS) operating system was installed in early June of this year on the IBM 370/168 to increase the efficiency of production processing.

Recommendation 5:

Commission a feasibility and cost-benefit study of a redesign and development in higher level language, of the business and individual master files and related applications, presently written in machine-dependent languages.

APPENDIX VI

Response:

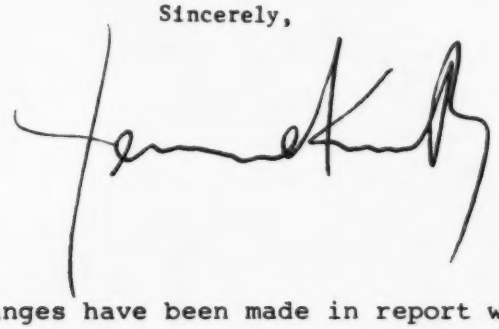
A decision to use higher level language for the Service's equipment replacement has already been reached as a result of the feasibility study for the Service Center Replacement System (SCRS). The Feasibility Study for the Master File Replacement System (MFRS) will be completed in April 1981 and will assume the use of COBOL (which is machine-independent). However, we cannot immediately undertake the redesign of the Master Files as suggested in the GAO Report (page 10) without seriously jeopardizing our work on SCRS and MFRS. Staff resources and skills are not available to do both simultaneously.

In addition to the above comments, we have noted in Enclosure A adjustments that need to be made to certain dates and figures in the draft GAO Report. (Enclosure deleted; see GAO note.)

Please let us know if we may be of further assistance in these matters.

With kind regards,

Sincerely,



Enclosure

GAO Note: Suggested changes have been made in report where appropriate.

(913650)